UNITED STATES PATENT AND TRADEMARK OFFICE

1-F)

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 10/648,177 | 08/26/2003 | Yifan Gong | TI-35984 | 4432 |

23494 7590 04/04/2007
TEXAS INSTRUMENTS INCORPORATED
P O BOX 655474, M/S 3999
DALLAS, TX 75265

| EXAMINER |
|---|
| SHAH, PARAS D |

| ART UNIT | PAPER NUMBER |
|---|---|
| 2609 | |

| SHORTENED STATUTORY PERIOD OF RESPONSE | MAIL DATE | DELIVERY MODE |
|---|---|---|
| 3 MONTHS | 04/04/2007 | PAPER |

**Please find below and/or attached an Office communication concerning this application or proceeding.**

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

| Office Action Summary | Application No. | Applicant(s) | |
|---|---|---|---|
| | 10/648,177 | GONG, YIFAN | |
| | Examiner | Art Unit | |
| | Paras Shah | 2609 | |

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on *26 August 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) *1-12* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-12* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☒ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *04 December 2003* is/are: a)☐ accepted or b)☒ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

## DETAILED ACTION

1. This communication is in response to the Application filed on 08/26/2003.

### *Drawings*

2. The drawings are objected to under 37 CFR 1.83(a). The drawings must show every feature of the invention specified in the claims. Therefore, the elements showing the steps of "out-of-vocabulary procedure" and "utterance acceptance as containing in-vocabulary word" must be shown or the feature(s) canceled from the claim(s). These can be incorporated in the form of a flow chart. No new matter should be entered.

Corrected drawing sheets in compliance with 37 CFR 1.121(d) are required in reply to the Office action to avoid abandonment of the application. Any amended replacement drawing sheet should include all of the figures appearing on the immediate prior version of the sheet, even if only one figure is being amended. The figure or figure number of an amended drawing should not be labeled as "amended." If a drawing figure is to be canceled, the appropriate figure must be removed from the replacement sheet, and where necessary, the remaining figures must be renumbered and appropriate changes made to the brief description of the several views of the drawings for consistency. Additional replacement sheets may be necessary to show the renumbering of the remaining figures. Each drawing sheet submitted after the filing date of an application must be labeled in the top margin as either "Replacement Sheet" or "New Sheet" pursuant to 37 CFR 1.121(d). If the changes are not accepted by the examiner,

the applicant will be notified and informed of any required corrective action in the next Office action. The objection to the drawings will not be held in abeyance.

### *Specification*

3. The disclosure is objected to because of the following informalities: “Appendix A” in paragraph [0022], line 4 should be “Appendix”.

Appropriate correction is required.

4. The disclosure is objected to because of the following informalities: “form” in paragraph [0030] should be “from”.

Appropriate correction is required.

5. The disclosure is objected to because of the following informalities: “fro” in paragraph [0063] should be “for”.

Appropriate correction is required.

### *Claim Objections*

6. Claims 1-12 are rejected to because of the following informalities:

As to claim 1, “the score difference” should be “a score difference” in line 5. Appropriate correction is required.

As to claim 4, “power density function” should be “probability density function” in line 1. Appropriate correction is required.

As to claim 5, a period needs to be added at the end of the claim.

As to claims 10-12, "the best possible log-likelihood" should be "a best possible log-likelihood" in line 3. Appropriate correction is required.

As to claim 10, "the cumulate log-likelihood" should be "the cumulate log-likelihood" in line 6. Appropriate correction is required.

As to claims 2,3, and 5-9 are objected to for being dependent on an objected to claim.

### *Claim Rejections - 35 USC § 112*

7. The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

8. Claims 1-12 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

9. Claim 1 recites the limitation "the recognized in-vocabulary" in line 6. There is insufficient antecedent basis for this limitation in the claim. Nonetheless, for the purposes of compact prosecution the limitation was interpreted as the commands in the vocabulary.

10. Claim 3 recites the limitation "the second section" in line 2. There is insufficient antecedent basis for this limitation in the claim. Nonetheless, for the purposes of compact prosecution the limitation was interpreted as the middle section.

11. Claim 4 recites the limitation "the two sections" in line 3. There is insufficient antecedent basis for this limitation in the claim. Nonetheless, for the purposes of

compact prosecution the limitation was interpreted as the first and last section for absorbing extra speech.

12. Claim 6 recites the limitation "the enrollment utterances" in line 2. There is insufficient antecedent basis for this limitation in the claim. Nonetheless, for the purposes of compact prosecution the limitation was interpreted as in-vocabulary words.

13. Claim 8 recites the limitation "the balance" in line 2. There is insufficient antecedent basis for this limitation in the claim. Nonetheless, for the purposes of compact prosecution the limitation was interpreted as error between a reference weight and the input weight for utterance.

14. Claim 9 recites the limitation “which has several alternative forms” in line 2-3. The preceding terms are indefinite as they claim many forms of a variety of elements in the claim, which is open ended. Nonetheless, for purposes of compact prosecution, a plurality of rejection parameters was interpreted. However, a suggestion of a plurality of rejection parameters can be stated.

15. Claim 10 recites the limitation "the first and last frames" in line 5. There is insufficient antecedent basis for this limitation in the claim. Nonetheless, for the purposes of compact prosecution the limitation was interpreted as the start and end of the command word.

16. Claim 11 recites the limitation "the first and last frames" in line 4. There is insufficient antecedent basis for this limitation in the claim. Nonetheless, for the purposes of compact prosecution the limitation was interpreted as the start and end of the command word.

17. Claim 11 recites "subtracting from a the above three values." There is insufficient antecedent basis for this limitation in the claim. It is unclear as to which of the three values the applicant is intending to use and from what the three values are being subtracted from as denoted by the "from a the" stated above. Nonetheless, for the purposes of compact prosecution the limitation was interpreted as the log-likelihood values of the non-speech being subtracted from the log-likelihood of the in-vocabulary word.

18. Claim 11 recites "the resulting value." There is insufficient antecedent basis for this limitation in the claim. It is unclear as to which resulting value the applicant is referring to. Nonetheless, for the purposes of compact prosecution the limitation was interpreted as result from the subtraction of the log-likelihood of the in-vocabulary word from the non-speech.

19. Claim 12 recites the limitation "the first and last frames" in line 4. There is insufficient antecedent basis for this limitation in the claim. Nonetheless, for the purposes of compact prosecution the limitation was interpreted as the start and end of the command word.

20. Claims 4, 5, 7, and 9 are rejected as being indefinite for being dependent upon an indefinite base claim.

### *Claim Rejections - 35 USC § 103*

21. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

22. Claim 1 is rejected under 35 U.S.C. 103(a) as being unpatentable over Lee *et al.* (US 6,519,563, issued on 02/11/2003) in view of Gupta *et al.* (US 5,390,278, issued on 02/14/1995).

As to claim 1, Lee *et al.* discloses a method for speaker-dependent voice command recognition comprising the steps of: providing a hybrid of phrases (see Abstract) (e.g. It is noted that a sentence network is a collection of phrases that form a sentence) and Gaussian mixture (see col. 8, line 27-36) (e.g. The reference denotes the use of Gaussian mixtures for each state; and performing a procedure to detect out of vocabulary words by calculating a difference between a top model (e.g. $\lambda_C$ is top model, which is referred to as "customer model") and a background model ($\lambda_B$ is referred to as the background model) (see col.3, equation in line 21-25 and) (e.g. It should be noted that the equation shown is the ratio of probabilities. The logarithm of the equation can be written as a log-likelihood difference between the model found and the background model, which is stated in the reference (col. 8- 9, line 66 and lines 1-4). Further, although the equation is shown in the background, the difference between the two log-likelihoods is used as mentioned by the calculation of the normalized score (see col. 8,

lines 63-67-col. 9 lines 1-4). However, Lee *et al.* does not specifically disclose the Gaussian mixture consisting of a pool of shared distribution. Gupta *et al.* does disclose the Gaussian mixture (see col. 6, lines 9-12) consisting of shared pool of distribution (see col. 6, lines 32-37) (e.g. It is noted that the covariance matrix is shared among nodes for the phoneme.) It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the method for voice command recognition presented by Lee *et al.* by the shared pool distribution teaching mentioned by Gupta *et al.* The motivation to combine the two references would involve the reduction of data size (see Gupta *et al.*, col. 6, line 32-33) in order to have reduced the amount of data to be analyzed by the system of Lee *et al.*

23. Claims 2 and 3 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lee *et al.* in view of Gupta *et al.* as applied to claim 1 above, and further in view of Wu ("Subsyllable-based discriminative segmental Bayesian network for Mandarin speech keyword spotting").

As to claim 2, Lee *et al.* and Gupta *et al.* do not specifically disclose wherein said network is a three section network. Wu does disclose wherein said network has three parts, where first and last sections are intended to absorb extra speech and the middle section to match in-vocabulary speech (see page 67, left column, lines 8-12 and Figure 3). It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have combined the speaker-dependent voice recognition presented by Lee *et al.* modified by Gupta *et al.* with the using of a three section network presented by Wu. The motivation to have combined the references involves the

extraction of non-keywords in order to determine keywords (see Wu, page 65, left column, lines 1-10). By ignoring the non-keywords, the system of Lee could have allowed the user to speak normally instead of only using allowable words (see Wu page 65, introduction, lines 7-10).

As to claim 3, Wu discloses wherein the first and last sections of a network comprise fully interconnected nodes and the second section comprises nodes sequentially connected (see Figure 3 and page 66, left column, lines 16-18) (e.g. It is seen from the figure that the left and right boxed elements depicting extraneous speech models and key word successor models are interconnected in a loop. The keyword model or the middle section is moving left to right. Also, of the first and last sections, the nodes are interconnected as seen from the latter citation. It is obvious to one skilled in the art that these models will consist of nodes. Further, the nodes are interpreted as being a grouping of network elements for a specific purpose. Since the reference deals with Bayesian networks, nodes for each of the models seen in Figure 3 are apparent and described in section 3.1, 1st paragraph).

24. Claims 4-6 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lee *et al. in* view of Gupta *et al.* in further view of Wu as applied to claim 3 above, and further in view of Newman *et al.* (US 6,151,575).

As to claim 4, neither Lee *et al.* nor Gupta *et al.* nor Wu specifically disclose each node having a PDF attached to the first and last sections and the sharing of the PDF for these sections. Newman *et al.* discloses the PDF being attached to a node and the sharing of PDF by common nodes (see col. 7, lines 33-42 and line 46). It would have

been obvious to one of ordinary skilled in the art at the invention was made to have modified the teachings presented by Lee and Gupta *et al.* and the use of a three section network presented by Wu with the use of a PDF being shared by each node. The motivation to have combined the references involves representing each node as a different model with similar properties being shared by nodes, which saves memory (see Newman *et al.* col. 7, lines 33-34 and line 47) for the system described by Lee *et al.* and the three different models described by Wu.

As to claim 5, Newman *et al.* discloses the PDF being modeled as a mixture of Gaussian distributions (see col. 7, line 12) (e.g. The single Gaussian distribution model can be also used from a mixture Gaussian distribution model by using a single distribution model rather than a sum of the models) with a unique variance shared by all nodes of the network (see col. 7 lines 38-39 and line 46) (e.g. Since silence segments or non-command segments contain similar models the use of the same model for specific nodes will allow the same variance to be shared as a result of the same PDF being used).

As to claim 6, Newman *et al.* discloses wherein the PDFs (see col. 7, lines 11-12) are trained (see col. 3, lines 62-67 and col. 7, lines 11-32) from the speech uttered (see Figure 2, elements 200 and 240) (e.g. In the Newman *et al.* reference the speaker-independent models consist of PDFs and thus are trained depending on the speaker models (see col. 7, lines 53-55). The PDFs trained from the second section would have been apparent as seen from the Wu reference, which describes a three section network

when using the teachings of Newman *et al.* Wu shows the training of keywords and non-keywords (see page 65, right column, 1st paragraph, lines 8-11)).

25. Claims 7-10, and 12 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lee in view of Gupta *et al.* in further view of Wu in further view of Newman *et al.* as applied to claim 6 above, and further in view of Lee *et al.* (US 5,675,706).

As to claim 7, neither Lee nor Gupta *et al.* nor Wu nor Newman *et al.* specifically disclose the first and last sections are the centroids of a clustering of the mean vectors of the second section. Lee *et al.* discloses the clustering of the word model vectors (see col. 3, lines 56-61) and finding the centroid (see col. 12, line 48 and col. 10, lines 1-3) of the result to form the anti-subword models (e.g. non-speech) (see col. 12, lines 11-14) (e.g. It is known in the art that the centroid is the mean vector. Thus, the determination of the centroid is done based on the clustering of the word model vectors (keywords or commands (e.g. in this case the second section (keyword) referred to by Wu)), which will form the PDF of the anti-subword discussed by Lee *et al.* (non-speech) (e.g. In the three section model presented by Wu, it is the keyword predecessor and keyword successor.) Further, the use of this reference allows the PDFs (discussed by Newman *et al.*) to be formed for the non-speech (first and last section) based on the clustering of the centroid of the speech as discussed in the Newman *et al.* reference. It should also be noted that the HMM models (see Lee *et al.* col. 6, lines 39-44) use the PDFs for determining the likelihood score). It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the teachings of Lee and Gupta *et al.* and Wu and Newman with the clustering of the word vectors to find the first

and last sections as presented by Lee *et al.* The motivation to have combined the references involve there being no assumption about the target keywords and non-keywords being made (see Lee *et al.* col. 2, lines 41-46) (e.g. Thus, it is not necessary for there to be training on the non-keywords). Further the use of this technique allows updates to the keywords to be performed (see Lee *et al.* (US 5,675,706) *et al.* col. 2, lines 61-64).

As to claim 8, Wu discloses wherein an adaptive weight (see page 67, left column, sect. 3.1, 1st paragraph, line 3) is attached to each node and the balance of recognition error (see page 67, left column, sect. 3.1, 1st paragraph, line 3 and lines 10-13) of the nodes of the network (e.g. Since the network consists nodes representative of Figure 3, the first and last sections are included. Further, the use of adaptive weights allows an implied adjustment to the weights due to recognition errors because the adaptive adjustment is commonly found to reduce a specific error between the reference template (trained speech values) and the input model (see page 67, left column, sect. 3.1, 1st paragraph, lines 11-13).

As to claim 9, Wu discloses a keyword spotting parameter (e.g. Similar concept to rejection hypothesis) for accepting an utterance based on al in-vocabulary word (see page 68, right column, sect. 3.3, equation 14 and last paragraph of section).

As to claims 10 and 12, Wu discloses wherein the rejection parameter is calculated using the following steps: calculating, the log-likelihood using a three section network model (see Figure 3), locating the first and last frame of the in-vocabulary word, extracting the cumulate log-likelihood from the first to the last frame of the in-vocabulary

word (see page 68, right column, equations 12, 13, and14) (e.g. It should be noted that the extraction part is inherent since the cumulate log-likelihood is calculated for the keyword (in-vocabulary word) and divided by the total time length as shown in equations 12 and 13 and the values are determined by the calculation of log-likelihoods over time.) calculating the best possible log-likelihood using a network model representing only the extra-speech from the first to the last frame of the in-vocabulary word (see page 68, right column, equations 12, 13, and 14) and dividing the difference of the above two values of log likelihood by the number of frames of the in-vocabulary word (see page 68, right column, equations 13 and14) (e.g. From the equation 13 it is evident that the log-likelihood is found for the keyword and background (garbage) models, which are then divided by the start and end times. Equation 14 represents a normalized ratio).

As to claim 11, Wu discloses wherein the rejection parameter is calculated using the following steps: calculating, the log-likelihood (see page 68, equation 12) (e.g. the formula for log-likelihood is given) using a three section network model (see Figure 3), locating the first and last frame of the in-vocabulary word (see page 68, right column, equations 13 and paragraph under equation) (e.g. It should be noted that the start and end times is dependent upon the lengths of the keywords trained (in-vocabulary word), calculating the best possible log-likelihood using a network model (see page 68, right column, equation 13) (e.g. It should noted that in equation 13 the log-likelihood of the keyword is found by using equation 12 and the accumulate log-likelihood is interpreted as the best log-likelihood.) representing only the extra-speech from the first to the last frame of the in-vocabulary word (see page 68, right column, equations 12, 13, and 14)

(e.g. The likelihood of the extra-speech is also found from equation 12 and 13) and dividing the resulting value (e.g. The resulting value was interpreted as stated above in the 35 USC 112 rejection of claim 11 (see bullet 17). The interpretation used was the subtraction of the log-likelihood of the in-vocabulary word from the extra-speech log-likelihood (e.g. Two log-likelihood values comprise the garbage (non-speech) model in the reference, which was stated as the keyword predecessor and keyword successor.) It is seen from equation 14 that the two likelihoods are subtracted by the reference.) by the number of frames of the in-vocabulary word (see page 68, right column, equations 13 and 14) (e.g. From the equation 13 it is evident that the log-likelihood is found for the keyword and background (garbage or non-speech) models, which are then divided by the start and end times. The end-time and start time depends upon the utterance length, which may be an in-vocabulary word or non-speech. Equation 14 represents a normalized ratio).

## ***Conclusion***

26. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Lennig (US 5,097,509), Arslan *et al.* (US 6,243,677) and Jiang *et al.* (US 6,502,072) recite a method for rejecting out of vocabulary utterances. Wilcox *et al.* (US 5,199,077), Vysotsky *et al.* (US 5,719,921), and Dharanipragada (US 6,073,095) recite a method for detecting spotting words in speech.

The NPL documents by Rose ("Discriminant wordspotting techniques for rejecting non-vocabulary utterances in unconstrained speech), Rose *et al.* ("Task independent wordspotting using decision tree based allophone clustering" ), Dharanipragada ("A fast vocabulary independent algorithm for spotting words in speech"), and Benayed ( "A new keyword spotting approach based on reward function") recite approaches to word-spotting for speech applications. The NPL documents by Ramalingam ("Speaker-dependent name dialing in a car environment with out-of-vocabulary rejection"), and Bazzi ("Modeling Out-of-vocabulary Words for Robust Speech Recognition") show method for modeling out-of-vocabulary rejection. The NPL document by Weintraub ("LVCSR log-likelihood ratio scoring for keyword spotting") is cited to teach a method for calculating log-likelihood ratio for keyword spotting.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Paras Shah whose telephone number is (571)270-1650. The examiner can normally be reached on MON.-FRI. 7:30a.m.-5:00p.m. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Xiao Wu can be reached on (571)272-7761. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

P.S.

03/19/2007

XIAO WU
SUPERVISORY PATENT EXAMINER